## Published as:

Rudnev M. Testing for invariance of latent classes: Group-as-covariate approach, in: Cross-cultural analysis: methods and applications, second edition. In E. Davidov, P. Schmidt, J. Billiet, B. Meuleman (eds.). London, Routledge, 2018. Ch. 16. P. 449-480. http://doi.org/10.4324/9781315537078

## Testing for invariance of latent classes: Group-as-covariate approach[1]

***Maksim Rudnev***

*National Research University Higher School of Economics*

### 1. Introduction

Development of classifications is a common task in the social sciences. Quite often the classifications are developed in a normative way, when group membership is assigned to a respondent or other unit of analysis based on some theoretical criteria or on the researcher's expectations (Doty & Glick, 1994). In other cases, classification is obtained empirically by the use of segmentation techniques such as hierarchical cluster analysis, *k*-means cluster analysis, or latent class analysis (LCA). Theoretical classifications are criticized for being speculative. In contrast, the empirical approach is frequently applied in a purely exploratory way, thus, the results might depend on the specific set of available data. Although LCA is vulnerable to such criticism to a lesser degree than the other classification techniques, it is frequently applied as an *ad hoc* segmentation solution. The validity and between-group reliability of empirical classifications are rarely tested.

At the same time, the tests for measurement invariance of factors, or continuous latent variables, have become a standard in cross-cultural studies (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014). Similar to comparison of results of the factor or any other measurement models, comparison of classifications across groups requires measurement invariance as well (McCutcheon, 1987, 2002; McCutcheon & Hagenaars, 1997). Although it is steadily spreading, assessment of invariance of classifications is still rarely applied (for applications, see Kankaraš, Moors, & Vermunt, and Siegers, this volume). This is partly due to the fact that classifications in general are less popular among quantitative social researchers as categorical latent variables require more sophisticated kinds of analysis, and computation of corresponding models is highly demanding in regard to the computing power, especially with larger samples and more complex models. And partly, this is because the methods for testing measurement invariance of classifications are less developed.

Kankaraš et al.'s contribution to this volume has presented a general overview of testing invariance with a large range of multiple group latent class models. This chapter focuses on one specific way of conducting such an analysis. We describe group-as-covariate approach, focus on

[1] This chapter is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). The author is grateful to the two anonymous reviewers and to the editors for their useful comments, which helped to significantly improve the chapter.

unordered latent class models, explicate levels of invariance and procedures required to test them making strong links with factor analysis, and supplement it with a detailed example. In addition to the application provided by Siegers (this volume), we describe and show how to test for metric invariance of classifications. In the following sections, we first describe a general LCA model, discuss how measurement invariance can be assessed, and finally demonstrate the application of the method using empirical data.

## 2. Measurement invariance of latent classes

### 2.1. Specifics of LCA in the context of measurement invariance

LCA is a family of models that look for a discrete latent variable using categorical observed variables (Goodman, 2002; McCutcheon, 1987). Typically, there is only one discrete latent variable, which might be viewed as the classification of the respondents based on their responses. Probabilities of these responses are set to be conditional on a discrete latent variable (unobserved classes) with some link function. In this chapter, we use the logistic link function, though there could be other link functions, for example, the probit function (Uebersax, 1999). The latent variable is always categorical, either nominal (this chapter) or ordered categorical (see Kankaraš et al., this volume).

In terms of probabilities (or probabilistic parameterization), there are three types of parameters in LCA: (1) the number of classes, (2) the probabilities of observed responses conditioned on class membership, these also may be labeled "class profiles",[2] and (3) the estimated sizes of the classes, or class probabilities. Any of these parameters can be constrained.

The purpose of assessing measurement invariance is to test whether measures are invariant across groups, that is, whether they have the same meaning. Conditional probabilities of responses are of central interest when assessing measurement invariance of classes across groups, because they define the substantive meaning of classes, thus giving them discriminative characteristics. The comparison of class sizes across groups makes sense only when conditional probabilities are (approximately) equal.

In the LCA context, we can discuss at least four invariance levels: no measurement invariance, configural invariance, full metric invariance, full scalar invariance, as well as partial scalar and metric invariance. Table 1 summarizes the levels of invariance and their implications. The number of classes might be less important as it does not always influence the meaning of all classes. For example, a researcher might be interested in comparing shares of heavy smokers between schoolboys and schoolgirls. In this case, it is not important whether, in each group, there

[2] It is not related to latent profile analysis, which is a generalization of the LCA model to continuous indicators (Vermunt, 2004). Class profile is a full set of conditional probabilities for the specific class.

are three or four classes of light smokers and nonsmokers defined by the same indicators. For this reason, one can discuss a certain level of invariance of a subset of classes if the invariance of the whole classification is not supported or not of interest.

*No invariance* of latent classes implies that some crucial part of the conditional probabilities of items are different across groups, and even a general configuration of item probabilities is not similar across groups. In other words, the meaning of latent classes is very different across groups; hence, the comparisons of the classes, its relations, and class proportions are not meaningful. *Configural invariance* is supported when a general configuration of conditional probabilities (probability of specific class member to give certain answer) is similar across groups but these probabilities are not necessarily equal. This points to the general similarity of class meaning across groups. *Metric invariance*[3] is difficult to define in terms of probabilities, but it implies, just as in factor analysis, that the relations of classes to the external variables are comparable across groups when conditional probabilities are not necessarily equal. *Scalar measurement invariance* requires equality of all conditional probabilities and implies the equality of class meanings, which allows researchers to correctly compare the class sizes. At the scalar level, the members of the similar classes are expected to give the responses to the set of indicator questions. The equality of conditional probabilities of the classes across groups implies that the respondents were classified with the same exact criteria in different groups. The *partial measurement* invariance is a situation in which some of the conditions of the full scalar or full metric invariance are not satisfied. The implications of partial invariance may be the same as for full invariance.

As conditional probabilities are not modeled directly (in a linear way), and a link function is involved, the levels of measurement invariance are defined differently from the ones reported for linear factor analysis with continuous latent variables. The next section elaborates it in detail.

### 2.2. Assessment of measurement invariance of classes with the group-as-covariate approach

The assessment of measurement invariance with confirmatory LCA involves modeling the LCA together with an observed group (e.g., country). Confirmatory LCA differs from an ordinary LCA simply by its higher number of constraints and the specific strategy to test nested models in order to justify or refuse a set of constraints (McCutcheon, 1987). This aspect is particularly useful in assessing the measurement invariance across groups.

[3] The label "metric" might be misleading in the context of LCA, as classes do not have a metric, still we keep it to compare with the levels of measurement invariance in factor analysis.

Probably the most developed approach is a multiple group LCA (Clogg & Goodman 1985; McCutcheon, 1987; Eid, Langeheine, & Diener, 2003; Kankaraš & Vermunt, 2014). Testing for measurement invariance involves estimation of the LCA model independently in each group, and then fitting another model where conditional probabilities are set to be equal across groups. The fit of the estimated models is then compared, and if the difference in fit indices is small and insignificant, a researcher can conclude that there is full scalar measurement invariance. In practice, thresholds are constrained rather than probabilities, as the former are the parameters in the model. This approach is equivalent to the group-as-covariate one, although the former is more flexible when testing for metric invariance and does not require proportional odds assumption, whereas the latter is more convenient because the group differences in thresholds are the actual model parameters rather than 'invisible' constraints. Analogously, this approach can be extended to the models with ordinal latent classes (see Kankaraš et al., this volume; Kankaraš, & Moors, 2009, 2012). There is evidence that ordinal LCA models are preferable over factor models in testing for measurement invariance. Simulation study by Kankaraš, Vermunt, and Moors (2011) demonstrated that factor analysis performs similarly when noninvariance is located in the slope parameters only, whereas LCA models detected it correctly. In this chapter, we focus on nominal latent class models that use ordinal indicator items.

Multilevel LCA is yet another approach where the thresholds are made random (i.e., allowed to differ across groups) and the models compared are (1) a model in which conditional probabilities are allowed to differ across groups (random thresholds) and (2) a model in which conditional probabilities are set to be equal across groups. Although this approach seems to be promising, especially when there are many groups, the applications are rare and the method itself is computationally heavy (see Henry & Muthén, 2010).

The most straightforward way to incorporate group into a LCA model is to supplement ordinary LCA with a categorical covariate that represents observed group membership. The grouping variable is allowed to affect each indicator, and these effects are different within each class, which might be thought of as, respectively, direct effect of group and interaction effect between class and group on observed indicators (McCutcheon & Hagenaars, 1997). We focus on this approach in the rest of the chapter.

The conditional probabilities are modeled through the threshold parameters, so roughly speaking, conditional probabilities in a simple logistic LCA are transformed thresholds. The general formulation of the response $j$ of the observed item $U_i$ given latent class $c$ and group $g$ as a covariate is the following:

$$\mathrm{P}(\mathrm{U}_i = \mathrm{j}|C = c, G = g) = \frac{\exp(\tau_{ijc} + \beta_{icg})}{1 + \exp(\tau_{ijc} + \beta_{icg})} \quad (1)$$

where $\tau_{ijc}$ is a threshold, specific for the category $j$ of indicator $i$ given a latent class $c$. The term $\beta_{icg}$ is an effect of group on conditional probability.[4] Note that $\beta_{icg}$ does not differ between categories $j$ within each indicator, so the effects of group on each of the categories within each indicator are assumed to be equal. It is referred to as proportional odds assumption and is necessary in the group-as-covariate LCA models. When $\beta_{icg}$ is zero, this implies the same expressions for each group, which is the invariance of this particular conditional probability. Therefore, the assessment of full scalar measurement invariance can be reduced to testing whether all $\beta_{icg}$ terms are equal to zero or not. In the context of factor analysis, the group-as-covariate approach (also known as MIMIC models) is limited to the assessment of scalar invariance only, as the covariate affects only the indicator intercepts. In the latent class model this is not the case, because the metric invariance model can be built by constraining group effects $\beta_{icg}$ to be the same across classes, but not necessarily equal to zero.

Model selection in this approach includes comparison of the model fit of the two models, where one of the models constrains all or some effects of group $\beta_{icg}$ and the other model does not constrain these effects. If the constrained model has a fit that is not substantially worse than the fit of the unconstrained model, a higher level of measurement invariance is supported. Therefore, testing measurement invariance involves a comparison of model fit between more constrained and less constrained models using statistical criteria. It is clear how to test full metric and scalar measurement invariance. However, most of the time this hypothesis is rejected, and the researcher has to decide between partial, configural, and no invariance. Due to the novelty of these methods and their relatively low popularity in the social sciences, statistical criteria like the comparative fit index (CFI) difference in factor analysis (Chen, 2007) or partial invariance rules (Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998) have, to the best of our knowledge, not been developed yet. Thus, in terms of guidance, a researcher is left with *ad hoc* solutions and intuitive substantive criteria in the selection between partial, configural, and no invariance.

The strategy in assessing the measurement invariance of LCA involves finding the optimal number of classes within each group, testing for full metric and scalar invariance, and, if full invariance is not supported, looking for and deciding between partial, configural, and no invariance. This procedure can be divided into the following six general steps.

[4] The formula is a simplified expression 13.5 given in the Kankaraš et al. chapter, this volume.

(1) Find an optimal number of classes within each group separately. This is a necessary and exploratory part of any LCA. The optimal number of classes involves estimation of 1-class, 2-class, and as many class models as can be considered feasible. Probably the most popular criteria of comparison are the Akaike and Bayesian information criteria (AIC and BIC, respectively); however, their differences do not have a known distribution, thus they cannot be compared using statistical tests. Comparison is performed following a simple rule—the smaller the BIC or AIC value, the better the model is.[5]

The likelihood ratio test (LRT), which is computed as a difference between –2*LogLikelihood values (or equivalent $G^2$ or $L^2$ statistics), may be used for comparison of the two nested models. However, the models with a differing number of classes are not truly nested. Nylund, Asparouhov, and Muthén (2007) demonstrated that the likelihood ratio, in a special case of testing a *k*-class model against a *k-1* class model, does not actually have a chi-square distribution. Thus, the LRT (as well as $G^2$ or $L^2$) cannot be used to detect the optimal number of classes. Some adjustments to this test were suggested, revealing Lo-Mendell-Rubin (Lo, Mendell, & Rubin, 2001) and Vuong-Lo-Mendell-Rubin LRTs, which were designed to detect differences in *k* and *k-1* class model fits. Bootstrapped LRT (McLachlan & Peel, 2000) is another approach to the LRT test. It predicts several datasets using estimated parameters from the *k-1* model and then analyzes them using *k-1* and *k* class models in order to approximate the distribution of –2*loglikelihood differences. The *p*-value produced by any LRT is (an approximation of) probability that the data were generated by the *k-1* model (Muthén & Muthén, 1998-2015). That is, a high *p*-value of a LRT indicates that a model with *k* classes should be rejected and a model with *k-1* classes should be accepted.

Alternative to the formal tests of model selection is a scaled entropy index, which reflects the reversed uncertainty of classification, that is, the certainty of classification (Asparouhov & Muthén, 2014). Values of the entropy index close to 1 indicate very high certainty of classification. Unlike the previous tests, entropy provides substantive information about classification in general. For example, Siegers (this volume) uses entropy over other criteria to detect the optimal number of classes.

It is not rare when information criteria, variants of LRT, and the entropy index contradict each other. Thus, the final decision on the number of classes is often made using substantive considerations, such as the similarity, interpretability, and meaningfulness of class profiles.

[5] Burnham and Anderson (2002) suggested rules of thumb for model selection for nested models using difference in AIC: 0-2 being negligible, 4-7 moderate, and more than 10 substantial. As models with different number of classes are not truly nested, it is not clear whether these rules are applicable here.

If there is the same number of classes in each group or a high level of invariance is expected, a top-down strategy is more feasible: It involves estimation of the full scalar invariant and noninvariant models (steps 2 and 5) and, if the comparison statistic is significant, then relaxing the constraints of the invariant model until acceptable fit is reached. In most other cases the down-top strategy is more feasible, that is, after detecting the number of classes, one can first explore the configural, metric, and kinds of partial invariance.

(2) Pool the data and build a single LCA model with a group added as covariate, allowing for group effects (no invariance assumed, sometimes referred to as the heterogeneity model, see Kankaraš et al., this volume). The only restriction is the equality of group effects within each indicator. At this stage one can inspect whether effects of group are significant, compute the share of significant effects, and compare these shares between classes and groups. It is indicative of similarity of classes across groups; higher numbers of insignificant effects indicate higher correspondence to full scalar invariance. If the share of insignificant group effects is large enough, it is reasonable to test for full scalar invariance. In other cases, configural and metric invariance should be tested first.

One of the possible ways for assessing configural invariance (Rudnev, Magun, & Schmidt, 2016) involves correlation of class profiles (sets of conditional probabilities). It involves calculation of class profiles, that is, full sets of conditional probabilities for each class, and then correlating them to each other. Squared Pearson correlation ($R^2$) might also be useful because it directly shows the percent of common variance of two profiles. If the correlations of class profiles are very high, it is enough to state configural invariance, and it is reasonable to test for higher levels of invariance.

(3) The metric invariance model implies the similar relations between a latent variable and its indicators across groups. These relations are defined here as distribution of the thresholds of the same indicator and the same category across classes. Metric invariance in LCA is similar to the one in CFA, because it assumes the similar strength of relations between latent variable (latent class) and its indicators across groups. Whereas in CFA the strength of these relations is represented by factor loadings, in LCA it is represented by differences in thresholds of the same indicator (and its category) across classes. In a general case, these relations are constrained to be equal separately for each category of each indicator, but in the group-as-covariate approach, where group effects are assumed to be the same within each indicator, it is reduced to constraining the group effects to be equal across all the classes. Note that it does not require group effects to be zero. Therefore, it makes the resulting differences between classes to be similar across groups, but does not require equality of conditional probabilities.

Like metric invariance in CFA, the support for this model allows between-group comparisons of within-group relations of classes with other variables. For example, the effect of gender on value classes in Eastern Europe and Western Europe can be compared if metric invariance criteria are met. However, metric invariance does not allow the comparison of class sizes across groups.

The metric invariance model is then compared to the unrestricted model. The fit statistics of these models, including a decrease in BIC and AIC, as well as LRT (which can be correctly applied in this situation) are then examined. In case the fit of these two models is not different, the constrained model is accepted as more parsimonious, and the metric measurement invariance of classes is supported. If metric invariance was supported, it is reasonable to test for full scalar invariance (step 5), and if not—there still might be partial metric invariance.

(4) There are two kinds of partial metric invariance—invariance of a subset of classes and invariance of a subset of indicators. In the first case, some classes hold metric invariance and others are not constrained. It requires at least two classes to be constrained because it is about differences between classes. In the second case, all the classes are constrained except for some indicators.

(5) Estimate a group-as-covariate LCA model where all the group effects are constrained to be zero or a model of full scalar invariance (also referred to as a structural homogeneity model). If this model's fit is not worse than that of the metric invariance model, one can accept the full scalar invariance. It implies that the class sizes can be meaningfully compared across groups and that predictors and outcomes of the latent classes can be added to the model and the resulting effects can be compared across groups. If full invariance is not supported, one may try to find partial invariance.

(6) Estimate another group-as-covariate LCA model with constraints defined by substantive criteria. For example, if there is a strong theory about one class, it might be tested for invariance whereas all other constraints may be relaxed. Or, if there is no theory at all, the group effects might be constrained stepwise based on the degree and significance of these effects in the unconstrained models. Note that partial scalar invariance is not compatible with partial metric invariance in LCA group-as-covariate approach because the partial scalar invariance assumes group effects to be zero for at least one class, whereas the partial metric model constrains them to be equal across classes, which results in a full scalar model. As noted earlier, very vague criteria exist for differentiating between configural and partial invariance. Results-led hypotheses for a model of partial scalar invariance might build up on the insignificant group effects. If a researcher is able to determine partial scalar invariance, the possibilities are open for comparing class sizes,

effects of its predictors, or outcomes across groups, although this should be done cautiously because strict statistical criteria is lacking.

**Table 1.** Levels of Latent Class Invariance, Requirements, and its Implications

| | **Configural invariance** | **Metric invariance** | **Scalar invariance** |
|---|---|---|---|
| Number of classes | Required* | Required* | Required* |
| Similarity of class profiles | Required | Required | Required |
| Similarity of differences in thresholds across classes | | Required | Required |
| Equality of thresholds across groups | | | Required |
| Allows comparison of… | signs of regression and correlations with external variables | regression and correlations with external variables; differences in class sizes. | regression and correlations with external variables; differences in class sizes; class sizes. |

**Note.** *Only when the researcher is interested in describing the whole population.
Partial metric and partial scalar invariance are met when part of the classes/indicators meet the criteria for full metric and scalar invariance.

## 3. Empirical example: European value classes in Western, Northern, and Eastern Europe

### 3.1. Substantive problem

Basic human values have been studied predominantly as continuous characteristics of people, however, both Schwartz (1992) and Rokeach (1973) emphasized the importance of the relevant position of values in a hierarchy of values. The typological approach brings to life the idea of analyzing value profiles instead of separate value characteristics. Latent value classes can differentiate between people with different individual hierarchies of values, which emphasizes the order of value preferences. Magun, Rudnev, and Schmidt (2016) have demonstrated that there are only five value classes in the European population, each of which has a substantially different value profile. The Growth class combines high importance of openness to change and self-transcendence values. The other four classes are aligned from a strong social focus (conservation and self-transcendence values) to a strong personal focus (self-enhancement and openness to change values). There are no classes that emphasize both self-enhancement and conservation. This typology remains stable across samples and time points (Rudnev et al., 2016).

This classification was obtained through the analysis of samples from the pooled European sample, putting together within- and between-country variance of values. Typically, a cross-cultural researcher is interested in finding measurement invariance between countries and, therefore, examines the within-country structure of the data. However, LCA is very sensitive to the amount of heterogeneity in the data, and some classes could be found only when there are respondents from several countries in the data. That is, some classes are cross-national by nature and thus hard to find within a national dataset. Magun et al. (2016) demonstrated that the most variable class across countries is the Growth values class—it has a higher probability in Western and Northern Europe and is hardly found in Eastern Europe. The other four classes seem to have less varying probabilities in all European countries (Magun et al., 2016).

In this paper, we assess the invariance of value classes in the three European regions as two groups: countries of Western and Northern Europe (hereafter designated as West&North Europe) comprise the first group and the countries of Eastern Europe comprise the second group. Previous analyses provide strong expectations that there are five value classes in West&North Europe, whereas there are only four of them (excluding the Growth values class) in Eastern Europe. The main hypothesis is that these four value classes, namely Strong Social Focus, Weak Social Focus, Weak Personal Focus, and Strong Personal Focus, possess measurement invariance between West&North and Eastern Europe.

### 3.2. Data

The data come from the European Social Survey (ESS). Round 4 was selected due to the highest number of participating countries. Overall, the sample includes 11 West&North European countries and 12 Eastern European countries, and specifically, 20,650 respondents are from West&North Europe and 21,854 from Eastern European countries. The countries are listed in the Supplementary materials. The values were measured by the 21 value portraits included in the ESS Human Values Scale (for the item wordings see Davidov et al., this volume, Table 6.2). The response options included six ordered categories ranging from "Very much like me" to "Not at all like me." The data were weighted with design and population weights because each group consists of several national samples and the countries differ in size.

### 3.3. Centering and method factor

In order to account for response style, or respondents' tendency to choose the same or similar ratings for the different value portraits, the use of centered instead of the raw ratings have been suggested (Schwartz, Verkasalo, Antonovsky, & Sagiv, 1997). Following these recommendations, within-individual centering is routinely conducted by the subtraction of the individual mean on all the ratings from each of the value ratings. We treated the six-point rating scale as a categorical response scale; thus, centering that involves averaging was not an option. Instead, the response style was modeled. Among many options available in LCA models for controlling for response style (see Kankaraš & Moors, 2011; Moors, 2003, 2004), the random effect model (Qu, Tan, & Kutner, 1996) appears to be the closest to the within-individual centering. It deals with the respondent's overall preference for a certain location of the response scale. Following the typology of response bias proposed by Van Vaerenbergh and Thomas (2013), both random intercept and within-individual centering capture acquiescence, disacquiescence, and midpoint response styles. The conventional LCA model, in addition to discrete latent variable (classes), was supplemented with a linear factor that represents response style (or random intercept, Vermunt, 2010)[6]. The loadings of this factor were constrained to 1 as response style has the same effect on all ratings irrespective of their content. The inclusion of this factor was necessary as the classes obtained without it reflected the response style *only*.

The resulting model looks like the model defined in expression (1) with a constrained factor *f* whose means can vary only across groups:

$$\mathrm{P}(\mathrm{U}_i = \mathrm{j} | C = c, G = g, f) = \frac{\exp(\tau_{ijc} + \beta_{icg} + \mathrm{f}_g)}{1 + \exp(\tau_{ijc} + \beta_{icg} + \mathrm{f}_g)} \quad (2),$$

where $\tau_{ijc}$ is a threshold, specific for the category *i* of indicator *j* given a latent class *c*; the term $\beta_{icg}$ is an effect of group on conditional probability; and $f_g$ is a random intercept whose mean differs across groups, but not across items or item categories.

### 3.4. General strategy of measurement invariance assessment

We do not expect the same number of classes in the two groups; therefore, we first test for the optimal number of classes, and second, we turn to full scalar invariance testing of the four classes. Next, if the scalar invariance is not supported, we look for configural and partial invariance of the classes.

All the models were fitted using Mplus 7.3 software (Muthén & Muthén, 1998-2015). The corresponding Mplus codes are listed in the Supplementary materials.

[6] Technically, such a model is called a factor mixture model (see Muthén, 2008).

Configural invariance is assessed by the correlation of full class profiles. As mentioned above, for this we computed conditional probabilities in the unconstrained model and stacked them by class and group and correlated them with each other. Pearson correlation of full class profiles indicates the similarity of distributions of response probabilities within indicators, as between-indicator variance does not exist in a full class profile. As an indicator of partial scalar invariance, we use the proportion of insignificant effects of a group variable in a nonconstrained group-as-covariate model. As the effects of group do not differ across categories within each indicator, the proportion of insignificant effects for each class would show the overall degree of invariance.

## 3.5. Results

### 3.5.1. How many classes are in different groups?

Determining the number of classes requires estimation of a large number of models. In each group we need to estimate a model with 1, 2, 3, and up to as many classes as it is feasible with the current data plus one. Thus, for example, with 2 groups and a hypothesized 5 classes we have to estimate at least 6 models in each group which results in 2 x 6 or 12 models. Given a large number of respondents in each group together with numeric integration required for this type of analysis, the estimation might involve a very long computation time. This is highly impractical. For this reason, each model was estimated in two steps (as recommended by Asparouhov & Muthén, 2012): First, the model with *k* classes was estimated without tests for *k-1* classes, and the optimal seed (a seed of random starts in maximum likelihood estimation) was found for the best likelihood value; and second, this seed was used to estimate both *k* and *k-1* class models. This approach significantly reduced computation time.[7]

Models with 1 to 7 classes have been estimated within each group. The results are listed in Table 2. Different model fit indices suggest differing optimal number of classes. Sample-adjusted BIC increases continuously from the model with just 1 class to the model with 7 classes, although the increase slows down with the higher number of classes. The entropy index is stable with different numbers of classes, with a value of around 0.79 in all class solutions for West&North Europe and around 0.84 for Eastern Europe. Although the entropy index tends to slightly decrease as the number of classes increases in Eastern Europe, it is hard to suggest the optimal number of classes based on these values. Thus, both the BIC and entropy indices are not helpful in suggesting

[7] This problem is sometimes referred as a "big data problem" in computer science and is often addressed by sampling of the data. Sampling procedure draws a random sample from each group of respondents, and then all the models required are estimated in a usual way using this smaller subsample. From our experience, application of sampling to LCA gives inaccurate results as compared to a full-sample estimation, which is due to the sensitivity of LCA to heterogeneity in the data.

the number of classes in our case. This may also signify the possibility that the respondents are not neatly clustered.

**Table 2.** Model Fit Statistics of Group-Specific Models with Different Numbers of Classes

| | **Entropy** | **Sample-adjusted BIC** | **–2*LogLikelihood** | **Number of parameters†** | **Lo-Mendell-Rubin LRT** | **Vuong-Lo-Mendell-Rubin LRT** |
|---|---|---|---|---|---|---|
| ***West&North Europe*** | | | | | | |
| 1 class | | 1,290,129 | 1,289,440 | 102 | | |
| 2 classes | 0.80 | 1,252,764 | 1,251,392 | 203 | 87,786* | 87,701* |
| 3 classes | 0.78 | 1,237,373 | 1,235,332 | 302 | 25,859* | 25,833* |
| 4 classes | 0.78 | 1,225,608 | 1,222,844 | 409 | 11,511* | 11,499* |
| 5 classes | 0.78 | 1,221,194 | 1,217,728 | 513 | 5,013* | 5,008* |
| 6 classes | 0.79 | 1,215,801 | 1,211,612 | 620 | 6,098* | 6,093* |
| 7 classes | 0.78 | 1,212,247 | 1,207,192 | 748 | 3,558 | 3,562 |
| ***Eastern Europe*** | | | | | | |
| 1 class | | 1,342,268 | 1,341,546 | 106 | | |
| 2 classes | 0.85 | 1,280,385 | 1,278,934 | 213 | 62,612* | 62,554* |
| 3 classes | 0.85 | 1,256,985 | 1,254,804 | 320 | 24,129* | 24,107* |
| 4 classes | 0.84 | 1,242,190 | 1,239,308 | 423 | 18,184* | 18,167* |
| 5 classes | 0.83 | 1,235,743 | 1,232,248 | 513 | 7,436 | 7,428 |
| 6 classes | 0.83 | 1,229,585 | 1,225,217 | 641 | 5,848 | 5,842 |
| 7 classes | 0.82 | 1,229,484 | 1,220,310 | 740 | 3,542 | 3,424 |

* *P*-value is lower than 0.05.

† In order to avoid indefinite parameters, some first or last thresholds were fixed to either -15 or 15 which translates into probability of 0 and 1, respectively.

The only available statistical tests of significance between models with differing number of classes are the two adjusted LRTs. Both the Lo-Mendell-Rubin and the adjusted Vuong-Lo-Mendell-Rubin LRTs provide the same results: The model fit significantly increases with the number of classes up to 7 classes in West&North Europe and 5 classes in Eastern Europe. It implies that the optimal number of classes in West&North Europe is 6 instead of the hypothesized 5, and 4 as expected in Eastern Europe. Since there were 5 classes hypothesized for West&North Europe, we compared the class profiles of the 5-class and 6-class solutions. This comparison reveals that the 6-class solution appears due to the splitting of the Weak Personal values class into two very similar classes. Full profile correlations between the Weak Personal values class in the 5-class solution with the two split classes in the 6-class solution are 0.96 and 0.86. Moreover, the predicted most likely membership of respondents is highly overlapped, dividing 85% of members of this class in the 5-class solution into the two in the 6-class solution. This finding indicates that it makes little substantive difference whether there are 5 or 6 classes in West&North Europe, and the 6-class solution is excessive. Hence, in contrast to the results of the LRT tests, we accept the 5-class solution on the grounds of substantive differentiation of classes.

#### 3.5.2. Is there scalar invariance of similar classes?

Since we found the expected number of classes, the next step is to take a shortcut and test the full scalar invariance against noninvariance. To do this, we pooled the data, and an LCA model was supplemented with a group variable, indicating whether the respondent is from West&North or from Eastern Europe. The diagram of this model is presented in Figure 1 and the formulation in (2).

In the noninvariant model (completely heterogeneous), the group variable is allowed to have an effect on every indicator within each class. In the fully scalar invariant model (structurally homogeneous), all the effects of group are set to zero, excluding the one on latent classes. As the optimal number of classes was found to be different in West&North and Eastern Europe, and the same discrete latent variable is used, the effect of group on one of the classes was constrained to be –15, as it translates to zero probability of this class in Eastern Europe. The Mplus codes for these models are listed in the Supplementary materials, Codes 1 and 2.

The model fit indices of the noninvariant (M0) model and model of a full scalar invariance (M3) are listed in Table 3. Scaled LRT[8] reveals that the difference in model fit of the noninvariant model is significantly better than that of the fully invariant model. This finding implies that we have to reject the full scalar invariance model and accept the noninvariant model.

However, there could still be metric invariance. Model M1 tests it against, again, the unrestricted model M0. LRT demonstrates a significant difference, and we have to accept the noninvariant model.

[8] Here, we use scaled LRT (Satorra & Bentler, 1999), as the models were estimated with the maximum likelihood robust algorithm which adjusts the likelihood for nonnormality.

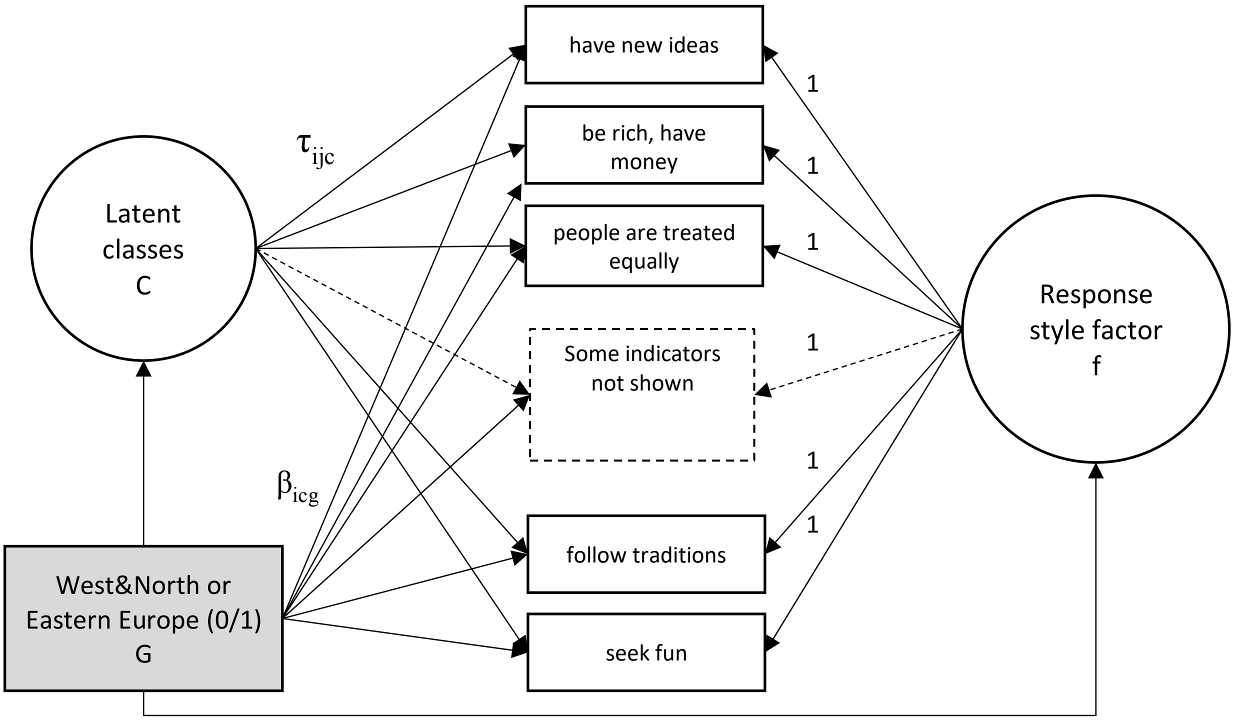


**Figure 1.** Group-as-covariate latent class model with "response style" factor. Eastern Europe is a dichotomous variable. Dotted arrows represent many other similar arrows.

### 3.5.3. Is there configural invariance between classes?

After rejecting the fully invariant models, the next step is to find an appropriate set of constraints which is satisfactory from both a statistical and a substantive standpoint. Table 4 lists the correlations of similar class profiles (for a full correlation matrix see Table 2 in the Supplementary materials) and the proportions of insignificant group effects from a common noninvariant model. In contrast to the rejection of the scalar invariance model, the correlations of the full class profiles are very high, ranging from 0.93 to 0.97. Figure 2 shows a high similarity of conditional probabilities obtained for West&North and East Europe. Nondiagonal correlations (i.e., between the other classes) are much lower, with the highest being 0.75 between Weak Personal and Weak Social classes, which is substantively justified given its shared "weakness" of profiles. In general, social values classes demonstrate higher similarity of class profiles.

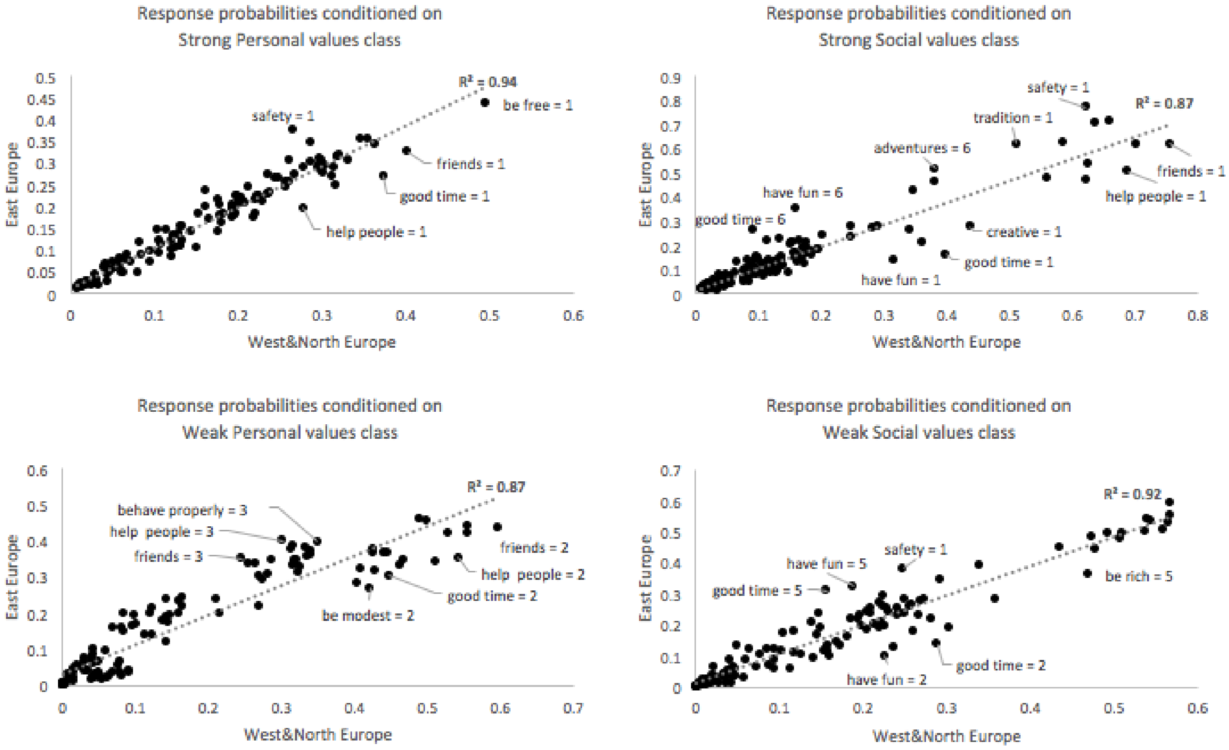


**Figure 2.** Scatterplots of conditional probabilities obtained in the two groups of countries in the unconstrained group-as-covariate LCA model M0.

Regarding the specific impact of the group on the level of conditional item probabilities, the group shows the smallest impact in the Weak Personal class profile, as the group has 16 (71%) insignificant effects out of a possible 21. Probabilities of items conditioned by Strong and Weak Social classes are more affected by the group, displaying only 5 and 4 insignificant group effects, respectively, and conditional probabilities within the Strong Personal class are fully dependent on the group. In general, only 30% of group effects are insignificant, which points to a low level of invariance. However, the significance of the group effects might be too strict, as correlations of value profiles demonstrate a high similarity of class structures.

Taking these similarities together, we defined two sorts of partially invariant models. The first was results-led and based on the estimates of model M0. Partially invariant metric model M2a constrains to equality only those group effects that have the smallest variance across classes, as it was estimated in model M0; those are most indicators excluding "equality," "behave properly," "environment," "tradition," and "safety." Analogously, a model of partial scalar invariance M4a is specified that fixes the smallest group effects to zero.

Taking into account the high correlations of full class profiles, it is hard to specify a partial metric invariance model for a subset of classes, so we simply took out the least invariant class in line with correlations, revealing model M2b. Making use of the fact that the Weak Personal class demonstrated the highest invariance across groups, M4b constrains its conditional probabilities to be equal across groups while relaxing all other parameters. M4c, in addition, constrains group effects on conditional probabilities of Weak Social class, and M4d constrains effects of group on conditional probabilities of Weak Personal, Weak Social, and Strong Social classes. All these models are nested in the noninvariant M0 model, so that they can be compared to M0 using LRT. The fit statistics and LRTs are listed in Table 3. All of the LRTs are significant, a finding which implies that the noninvariant model M0 has to be accepted. However, again, these *p*-values may be overly sensitive to substantively unimportant differences in conditional probabilities and/or due to a large sample size. Although not a subject of statistical significance, adjusted BIC indices may help in finding an optimal solution. Given the very large BIC value in the noninvariant model M0, the growth of BIC in the more constrained models is very small: For partial metric models it is 0.03% and 0.04%, and for the scalar invariance M5 it barely reaches 0.40%. When using this criterion, even the scalar invariance model can be accepted.

**Table 3.** Model Fit Statistics for the Fitted Latent Class Analysis Models

| | –2*Log Likelihood | Number of parameters | Scaled LRT | Sample-adjusted BIC | Difference of adjusted BIC with M0 |
|---|---|---|---|---|---|
| **M0.** Noninvariant | 3,140,797 | 622 | | 3,145,616 | |
| **M1.** Full metric invariance | 3,142,746 | 559 | 579.2 (63)* | 3,147,076 | 1,460 (0.05%) |
| **M2a.** Partial metric invariance (some indicators are not invariant, results-led) | 3,142,125 | 574 | 391.5 (48)* | 3,146,571 | 955 (0.03%) |
| **M2b.** Partial metric invariance of Weak Personal, Weak Social, and Strong Social classes | 3,142,221 | 580 | 444.5 (42)* | 3,146,714 | 1,098 (0.04%) |
| **M3.** Full scalar invariance | 3,153,992 | 538 | 4,348.5 (80)* | 3,158,159 | 12,544 (0.40%) |
| **M4a.** Partial scalar invariance: Some indicators are not invariant, results-led | 3,141,434 | 595 | 245.6 (27)* | 3,146,043 | 428 (0.01%) |
| **M4b.** Scalar invariance of Weak Personal class is invariant | 3,142,604 | 598 | 691.1 (24)* | 3,147,237 | 1,621 (0.05%) |
| **M4c.** Scalar invariance of Weak Personal and Weak Social classes are invariant | 3,146,358 | 577 | 3,250.1 (45)* | 3,150,827 | 5,212 (0.17%) |
| **M4d.** Scalar invariance of Weak Personal, Weak Social, and Strong Social classes are invariant | 3,151,149 | 556 | 4,149.8 (66)* | 3,155,456 | 9,841 (0.31%) |

Note. * *P*-value is less than 0.05. The number of parameters consists of 21 indicators with five thresholds each, variance of method factor differing across four classes, three class sizes thresholds, and all these parameters estimated in two groups. Some of the thresholds were automatically fixed during estimation in order to avoid infinity.

**Table 4.** Correlations of Full Class Profiles in West&North Europe and Eastern Europe and the Proportions of Insignificant Effects of Group on Conditional Probabilities of Items

| **Class label** | **Correlations of full profiles,** $N = 126$ | **% of insignificant at** $p > 0.001$ **effects of group,** $N = 21$ |
|---|---|---|
| Growth | - | - |
| Strong Social | 0.93 | 23% (5) |
| Weak Social | 0.96 | 19% (4) |
| Weak Personal | 0.94 | 71% (16) |
| Strong Personal | 0.97 | 0% (0) |

**Note.** The probabilities for correlations and the number of insignificant effects are taken from model M0, model fit listed in Table 3. All correlations are significant at $p < 0.01$.

At this point we have to make a decision about the level of measurement invariance of the four classes in West&North and Eastern Europe. Instead of reaching a single conclusion, we take into account all the information and suggest that it depends on the strictness of the criteria applied. Following the strictest criteria of the current literature, configural invariance requires the same number of classes in all groups, so one may say the current value classes do not reach even configural invariance, because we ended up with different number of classes. However, as discussed in section 2.1, we suggest to apply the term "invariance" in the LCA context to the specific subsets of classes rather than to the whole classifications. Thus, we can say that there is configural measurement invariance of the four value classes across two groups of countries, and this is concluded from the rejection of all partially invariant models, taken together with the high correlations of class profiles. It implies that the classes have the same structure across groups, but the class sizes cannot be compared. If the LRT $p$-values are considered too strict, and the BIC a too weak criterion, the conclusion is subject to a researcher's judgment. Using correlations of value profiles, the partial invariance of the Weak Social, Weak Personal, and Strong Social classes may be stated, whereas the Strong Personal class is not invariant, as all of its conditional probabilities are affected by the group. This implies that only the three classes can be compared across groups. Table 5 reports the class probabilities as estimated in model M4d. However, to compare class probabilities, one more step might be needed—because the Growth class was found to be nonexistent in Eastern Europe, class probabilities were recalculated without the Growth class in West&North Europe. The recalculated probabilities of Strong Social and Weak Personal classes are similar in the two groups, however, there is a large difference in the probability of Weak Social class, which is two times higher in Eastern than in West&North Europe. This finding can be taken as evidence that East Europeans are more prone to Weak Social orientation than the inhabitants of West&North Europe.

**Table 5.** Estimated Latent Class Probabilities in West&North and Eastern Europe

| | **West&North Europe** | **West&North Europe, recalculated excluding the Growth class** | **Eastern Europe** |
|---|---|---|---|
| Growth | 24% | - | - |
| Strong Social | 15% | 20% | 18% |
| Weak Social | 13% | 17% | 31% |
| Weak Personal | 20% | 27% | 24% |
| Strong Personal (uncomparable) | 27% | 36% | 27% |

## 4. Conclusion

In this chapter we described and demonstrated one way to test for measurement invariance of latent classes. This approach involved looking for the number of classes within each group, estimating a single model with a group-as-covariate observed variable, and applying different sets of constraints to this model in order to test full and different kinds of partial invariance. Correlation of full class profiles were suggested as a way to assess configural invariance. We demonstrated this approach using the data on basic human values in two groups of countries—West&North Europe and Eastern Europe. The formal criteria pointed to the existence of only configural invariance of the four latent classes across the two groups. However, these criteria might be too strict and, given the substantial considerations, it might be concluded that the three latent classes are invariant across the groups. Using this result, the class probabilities were legitimately compared, and analyses revealed Eastern Europeans were much more prone to membership in the Weak Social values class than members of the West&North European population.

Overall, further research is needed to guide the model selection process for deciding between partial and configural measurement invariance of latent classes as well as to develop model fit statistics that are more flexible and robust to a sample size, alternative to the overly conservative likelihood ratio test. The overly strict LRTs and uncertainty of information criteria such as BIC need to be amended with newer approaches, such as newly introduced approximate

(Bayesian) measurement invariance in factor analysis. Nonetheless, there are several tools currently available to test for measurement invariance of latent classes across groups, some of which were demonstrated in this chapter.

## References


Asparouhov, T., & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. Mplus Web Notes: No. 14. May 22, 2012. Retrieved from https://www.statmodel.com/examples/webnotes/webnote14.pdf

Asparouhov, T., & Muthén, B. (2014). Variable-specific entropy contribution. Mplus Web Notes. October 7, 2014. Retrieved from https://www.statmodel.com/download/UnivariateEntropy.pdf

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin,* 105(3), 456-466. doi:10.1037/0033-2909.105.3.456

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504. doi:10.1080/10705510701301834

Clogg, C. C., & Goodman, L. A. (1985). Simultaneous latent structure analysis in several groups. In N. B. Tuma (Ed.), *Sociological Methodology* (pp. 81-110). San Francisco, CA: Jossey-Bass Publishers.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*, 55-75. doi:10.1146/annurev-soc-071913-043137

Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, *19*(2), 230-251. Retrieved from http://www.jstor.org/stable/258704

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis a primer. *Journal of Cross-Cultural Psychology*, *34*(2), 195-210. doi:10.1177/0022022102250427

Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 3-55). Cambridge, United Kingdom: Cambridge University Press.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193-215. doi:10.1080/10705511003659342

Kankaraš, M., & Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe insights from a multiple-group latent-class factor approach. *International Sociology, 24*(4), 557-579. doi:10.1177/0268580909334502

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across Europe. *Methodology, 7,* 68-80. doi:10.1027/1614-2241/a000024

Kankaraš, M., & Moors, G. (2012). Cross-national and cross-ethnic differences in attitudes: A case of Luxembourg. *Cross-Cultural Research, 46*(3), 224-254. doi:10.1177/1069397112440945

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research, 40*(2), 279-310. doi:10.1177/0049124111405301

Kankaraš, M., & Vermunt, J. K. (2014). Simultaneous latent-class analysis across groups. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 5969-5974). Dordrecht, Netherlands: Springer.

Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity, 37*(3), 277-302. doi:10.1023/A:1024472110002

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review, 20*(4), 303-320. doi:10.1093/esr/jch026

Qu, Y., Tan, M., & Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics, 52*(3), 797-810. doi:10.2307/2533043

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767-778. doi:10.1093/biomet/88.3.767

Magun, V., Rudnev, M., & Schmidt, P. (2016). Within- and between-country value diversity in Europe: A typological approach. *European Sociological Review, 32*(2), 189-202. doi:10.1093/esr/jcv080

McCutcheon, A. L. (1987). *Latent class analysis*. Newberry Park, CA: Sage Publications.

McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 56-88). Cambridge, United Kingdom: Cambridge University Press.

McCutcheon, A. L., & Hagenaars, J. A. (1997). Comparative social research with multi-sample latent class models. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 266-277). Münster, Germany: Waxmann.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1-24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov. T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535-569. doi:10.1080/10705510701575396

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Rudnev, M., Magun, V., & Schmidt, P. (2016). Basic human values: Stability of value typology in Europe. In M. Voicu, I. C. Mochmann, & H. Duelmer (Eds.), *Values, economic crisis and democracy* (pp. 21-49). Abingdon, United Kingdom: Routledge.

Satorra, A., & Bentler, P. (1999). *A scaled difference chi-square test statistic for moment structure analysis*. Universitat Pompeu Fabra, Department of Economics Working Paper No. 412. Retrieved from https://ssrn.com/abstract=199064

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). Orlando, FL: Academic Press.

Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology, 36*(1), 3-18.

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90. doi:10.1086/209528

Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement, 23*(4), 283-297. doi:10.1177/01466219922031400

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*(2), 195-217. doi:10.1093/ijpor/eds021

Vermunt, J. K. (2004). Latent profile model. In M. S. Lewis-Beck, A. Bryman, & T. Futing Liao (Eds.), *The Sage encyclopedia of social sciences research methods* (pp. 554-555). Thousand Oaks, CA: Sage Publications, Inc.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450-469. doi:10.1093/pan/mpq025

# Supplementary materials

**Table 1.** The list of countries included in the analysis.

| **West&North Europe** | **East Europe** |
|---|---|
| Belgium | Bulgaria |
| Denmark | Croatia |
| Finland | Czech Republic |
| France | Estonia |
| Germany | Hungary |
| Ireland | Latvia |
| Netherlands | Poland |
| Norway | Romania |
| Sweden | Russia |
| Switzerland | Slovakia |
| United Kingdom | Slovenia |
| | Ukraine |

**Table 2.** Full class profiles correlations (based on estimated condition item probabilities in the unconstrained model M0)

| | East Europe | | | |
|---|---|---|---|---|
| **West&North** | **Strong Personal** | **Weak Personal** | **Weak Social** | **Strong Social** |
| **Growth** | 0.59** | 0.55** | 0.59** | 0.14 |
| **Strong Personal** | 0.**97**** | 0.52** | 0.47** | 0.31** |
| **Weak Personal** | 0.68** | **0.94**** | 0.75** | -0.11 |
| **Weak Social** | 0.57** | 0.67** | 0.**96**** | 0.28** |
| **Strong Social** | 0.41** | -0.25** | 0.22* | **0.93**** |

* *P*-value is lower than 0.05.
** *P*-value is lower than 0.01.

**Code 1. Annotated Mplus code for the unconstrained group-as-covariate model (M0)**

```
DATA:
    FILE IS "tabdelimited.dat";
    LISTWISE=off;
VARIABLE:
    NAMES ARE  idno v1-v21 pweight East;
    MISSING are ALL (999);
    USEVARIABLES =  v1-v21 East;
    WEIGHT= pweight;  ! applies population weight. Exclamation mark stands for comment
    CATEGORICAL = ALL;
    CLASSES = KLUSTER(5);    ! this is the name of latent class variable and the number of classes in both groups which must be equal

ANALYSIS:
    TYPE = MIXTURE;  ! designates the type of analysis that looks for latent classes
    ESTIMATOR = MLR;
    STARTS = 50 10;
    ALGORITHM = INTEGRATION;
    PROCESSORS=5;    ! this takes advantage of several cores of processor in order to endorse computation speed

 MODEL:
%overall%
         v1-v21 ON East*0;  ! the effect of group variable East on conditional probabilities of observed indicators can be any value, though the starting value is zero. These are βicg coefficients.
        kluster#1 ON East@-15;  !  this fixes the probability of class "kluster#1" to zero by fixing the effect of the group variable to a value of -15
        kluster#2 ON East;
        kluster#3 ON East;
        kluster#4 ON East;  ! these lines indicate that three classes (as well as referenced class "kluster#5") can be affected by group variable East.

        method_f BY v1-v21@1;    !  this is the method factor used to adjust for the response style, it has loadings on all the 21 items and all of them are fixed at 1.
         [method_f@0];  !the mean of the method factor is fixed to 0
         method_f ON East*; !the method factor can differ between groups
```

```
%kluster#1%
    [v1$1-v21$5];
```

! this designates that the thresholds of indicators in class "kluster#1" can differ from the ones in the other classes. These are $\tau_{ijc}$ coefficients, namely $\tau_{ij1}$

```
    v1-v21 ON East@0;
```

! the effect of group variable East on conditional probabilities of observed indicators is fixed to zero. These are $\beta_{icg}$ coefficients, namely $\beta_{i1g}$.

```
    method_f;
```

! this designates that method factor can have different variance across classes

```
    method_f ON East@0;
```

! the effect of group on method factor needs to be fixed for the "kluster#1" class because this class is fixed to zero at group "East"

```
%kluster#2%
    [v1$1-v21$5*];
    v1-v21 ON East*0 (c2_i1-c2_i21);
```

! the effect of group variable East. These are $\beta_{icg}$ coefficients, namely $\beta_{i2g}$. In parentheses labels of parameters are given that will be used in MODEL CONSTRAINTS section. c stands for class id, i stands for indicator id

```
    method_f;
```

! the code for the remaining classes is the same as for the class "kluster#2" except for parameter labels

```
%kluster#3%
     [v1$1-v21$5*];
    v1-v21 ON East*0(c3_i1-c3_i21);
    method_f;

%kluster#4%
     [v1$1-v21$5*];
    v1-v21 ON East*0 (c4_i1-c4_i21);
    method_f;

%kluster#5%
     [v1$1-v21$5*];
    v1-v21 ON East*0(c5_i1-c5_i21);
    method_f;
```

**Code 2. Annotated Mplus code for the full and partial invariance models M1-M4d**

These codes are different from Code 1 only by adding MODEL CONSTRAINT statements.

DO command repeats the command replacing character # with numbers sequence from the first in parentheses to the last in parentheses.

**Metric invariance model M1:**

```
MODEL CONSTRAINT:
      DO(1,21) 0 = c3_i# - c2_i#;
      DO(1,21) 0 = c4_i# - c3_i#;
      DO(1,21) 0 = c5_i# - c4_i#;
```

**Partial metric invariance model M2a where several indicators are not constrained:**

```
MODEL CONSTRAINT:
      DO(3,5) 0 = c#_i1 - c2_i1;
      DO(3,5) 0 = c#_i2 - c2_i2;
      !DO(3,5) 0 = c#_i3 - c2_i3;
      DO(3,5) 0 = c#_i4 - c2_i4;
      DO(3,5) 0 = c#_i5 - c2_i5;
      DO(3,5) 0 = c#_i6 - c2_i6;
      DO(3,5) 0 = c#_i7 - c2_i7;
      DO(3,5) 0 = c#_i8 - c2_i8;
      DO(3,5) 0 = c#_i9 - c2_i9;
      DO(3,5) 0 = c#_i10 - c2_i10;
      DO(3,5) 0 = c#_i11 - c2_i11;
      DO(3,5) 0 = c#_i12 - c2_i12;
      DO(3,5) 0 = c#_i13 - c2_i13;
      !DO(3,5) 0 = c#_i14 - c2_i14;
      DO(3,5) 0 = c#_i15 - c2_i15;
      !DO(3,5) 0 = c#_i16 - c2_i16;
      DO(3,5) 0 = c#_i17 - c2_i17;
      DO(3,5) 0 = c#_i18 - c2_i18;
      !DO(3,5) 0 = c#_i19 - c2_i19;
      !DO(3,5) 0 = c#_i20 - c2_i20;
```

```
DO(3,5) 0 = c#_i21 - c2_i21;
```

**Partial metric invariance model M2b where class 2 is not constrained, but group effects are constrained to be equal across classes 3, 4, and 5:**

```
MODEL CONSTRAINT:
      !DO(1,21) 0 = c3_i# - k2_i#;
      DO(1,21) 0 = c4_i# - c3_i#;
      DO(1,21) 0 = c5_i# - c4_i#;
```

**Full scalar invariance model M3**

```
MODEL CONSTRAINT:
      DO(1,21) 0 = c2_i#;
      DO(1,21) 0 = c3_i#;
      DO(1,21) 0 = c4_i#;
      DO(1,21) 0 = c5_i#;
```

**Partial scalar invariance model M4a**

```
MODEL CONSTRAINT:
      0 = c3_i1;
      0 = c3_i3;
      0 = c3_i4;
      DO(6,12) 0 = c3_i#;
      0 = c3_i16;
      DO(18,21) 0 = c3_i#;

      0 = c4_i10;
      0 = c4_i12;
      0 = c4_i18;

      0 = c5_i6;
      0 = c5_i8;
      0 = c5_i12;
```

**Partial scalar invariance model M4b**

```
MODEL CONSTRAINT:
      !DO(1,21) 0 = c2_i#;
      DO(1,21) 0 = c3_i#;
      !DO(1,21) 0 = c4_i#;
      !DO(1,21) 0 = c5_i#;
```

**Partial scalar invariance model M4c**

```
MODEL CONSTRAINT:
      !DO(1,21) 0 = c2_i#;
      DO(1,21) 0 = c3_i#;
      DO(1,21) 0 = c4_i#;
      !DO(1,21) 0 = c5_i#;
```

**Partial scalar invariance model M4d**

```
MODEL CONSTRAINT:
      !DO(1,21) 0 = c2_i#;
      DO(1,21) 0 = c3_i#;
      DO(1,21) 0 = c4_i#;
      DO(1,21) 0 = c5_i#;
```